1958

An Analysis of Item Order in Seven Subtests of the Wechsler Adult Intelligence Scale

William Gregory Klett
Loyola University Chicago

AN ANALYSIS OF ITEM ORDER IN SEVEN SUBTESTS OF THE WECHSLER ADULT INTELLIGENCE SCALE

by

William G. Klett

A Thesis Submitted to the Faculty of the Graduate School of Loyola University in Partial Fulfillment of the Requirements for the Degree of Master of Arts

February

1958

LIFE

William Gregory Klett was born in St. Paul, Minnesota, December 9, 1929. He was graduated from Cretin High School, St. Paul, Minnesota, June, 1947, and from The College of St. Thomas, also in St. Paul, June, 1951, with the degree of Bachelor of Arts.

From July, 1952 to July, 1954, the author served as a personnel management specialist for the 9829 Technical Service Unit, United States Army Corps of Engineers. While in the service he attended George Washington University and The United States Department of Agriculture Graduate School, both of Washington, D.C.

Following his honorable separation from the United States Army he was employed as a merchandiser and auditor by Montgomery Ward and Company. He began his graduate studies at Loyola University in February, 1955.

In June, 1956, he resigned from Montgomery Ward and accepted the position of psychologist at the Chicago Catholic Charities, Child Guidance Clinic.

ACKNOWLEDGMENTS

The writer is indebted to Frank J. Kobler, Ph.D., Professor of Clinical Psychology, Loyola University, under whose direction this study was completed, for his interest and helpful criticism. Grateful acknowledgment is also made to Ralph D. Bergen, M.D., Director of the Catholic Charities Guidance Center and to Fred D. Whelan, Ph.D., for their encouragement and personal assistance. The writer also wishes to express his appreciation to Mrs. Joan Maestranzi and Mrs. Marlene Klett for their assistance in the preparation of this manuscript.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

STATEMENT OF THE PROBLEM

The Wechsler-Bellevue Intelligence Scale, Form I, was introduced in 1939 to fulfill the well-recognized need for an individual intelligence test suitable for adults.[1] Existing intelligence tests, notably the Stanford-Binet Scale, had been found lacking in interest and of insufficient discriminative value for older subjects. For the most part, the available tests were age scales that had been standardized on school children. In view of the inapplicability of mental age norms to adults and other limitations inherent in the tests at hand, Wechsler developed a point scale consisting of six verbal and five performance type subtests. A comparable scale, Form II, was developed for military use during World War II.

In the eighteen years since its introduction, the W-B I has been the object of numerous experimental investigations and critical evaluations. These studies revealed certain deficiencies and limitations in the test's construction, standardization, diagnostic utility, and administration procedure. The principal criticisms were directed against its normative population, an all

---

[1]Hereafter referred to as W-B I.

white sample from the state of New York; the reliabilities of certain subtests; and its restricted range of item difficulty.

Several item analyses of the W-B I indicated that the items of some subtests were not arranged according to difficulty value, i.e., they were not in order of least to most difficult. Rabin et al has periodically reviewed and summarized the reported W-B studies (8,16,18).

It was to correct these deficiencies and limitations that the Wechsler Adult Intelligence Scale[2] was developed as an extension and modification of the W-B I and introduced in February, 1955. A representative sample of the United States population was employed in standardizing the WAIS and numerous changes in administration and scoring procedure were made. Ambiguous items have been replaced and new ones added. The revised scale contains 257 items whereas the original one has 217 items. This increase in number of items is an attempt to provide the WAIS with an adequate range of item difficulty. The number of items retained from the W-B I and the new ones added are indicated in Table IX of the Appendix. A comparison of the W-B I and WAIS record blanks indicates that with respect to each other the positions of many retained items have been changed, apparently because their difficulty values have fluctuated since 1939. These positional changes may also reflect the more extensive and representative procedures employed in standardizing the WAIS.

It is the purpose of this investigation to explore the present item order of the Information, Comprehension, Arithmetic, Similarities, Vocabulary, and

---

[2]Hereafter referred to as WAIS.

Block Design, and Picture Arrangement subtests of the WAIS in order to determine their applicability to an eighteen and nineteen year old sample of males and females. This purpose is of value because instructions for each of the WAIS subtests under study permit the examiner to terminate testing when a prescribed number of consecutive failures (responses scored zero) have been recorded. In view of this allowance, it is possible that examiners may fail to administer items that are within the range of a testee's abilities if they are not arranged according to difficulty value. Also, if WAIS items are incorrectly ordered, a more accurate positioning of items would facilitate economy of time and greater ease in testing. Finally, stable and well-graded item orders are essential to the derivation of clinically meaningful interpretations of interest scatter.

CHAPTER II

RELATED LITERATURE

As far as can be determined from a review of the literature, no investigation of the item orders of the WAIS subtests has been published. Several item analyses of the W-B I and similar studies of the Revised Stanford-Binet Intelligence Scale have been reported.

Altus (26, p. 172) found marked differences in the difficulty value of items in the Information subtest of the W-B I. As a result of his item analysis changes in the item order of the Information subtest were made by Wechsler in the third edition of the W-B I. Altus' revised item order was based on the responses of two hundred, randomly selected draftees. The results were item analyzed by the Q_4-Q_1 technique to establish the internal validity of each of the items. Correct responses were then counted and divided by two to obtain the percentage difficulty value of each item. In a personal communication to this writer Altus wrote, "I suspect that information items in any test should be recalibrated in difficulty value about every five years. For instance, the item, "Where is Tokyo?" would have been much less difficult ten or twelve years ago than it is now, but the validity would probably have been relatively constant."[1] Unfortunately, Altus restricted his analysis to one subtest. In view of the

---

[1] Altus, W. D. Personal Communication, 1957.

marked differences in difficulty value of the Information items that he found, it would have been advisable to investigate the item orders of other W-B I subtests. Also, Altus stated in his communication that the Information subtest was administered at a Los Angeles, California, induction station. Hence, his revised order was derived from the test records of an all male sample, most of whom were probably born and raised in California. On the other hand, Wechsler's original order was based on the responses of a New York State sample. One would expect the difficulty value of certain Information items to vary when administered to samples of heterogenous populations living approximately 3,000 miles apart. Finally, it was disconcerting to note that Altus' study was not published so that a more thorough analysis of the data could be made.

Rapaport (19) likewise studied the W-B I and found orders of item difficulty which differed from those of Wechsler's. He presented an analysis of the subtest items to establish whether success or failure on single items can be diagnostic, i.e., to discover whether there are individual items on which failure is characteristic for any clinical group. His sample included patrolmen and variously diagnosed psychiatric patients. Rapaport divided subtest items into groupings of various degrees of difficulty, e.g. "easiest," "more difficult," and "most difficult," but did not calculate rank order correlations between his item orders and those of Wechsler. He commented that it was common to find marked deviations from the item orders suggested by Wechsler. The principal value of Rapaport's investigation, in so far as this study is concerned, was that it pointed out these deviations in several of the W-B I subtests and cautioned against accepting Wechsler's item arrangements at

their face value. It should be noted, however, that Rapaport's sample of patrolmen and variously diagnosed psychiatric patients was not representative of the United States population.

In a study more directly related to the topic of this paper, Rabin, Davis, and Sanderson (17) reported the order of difficulty for the W-B I items of the Information, Comprehension, Similarities, Picture Arrangement, Picture Completion, and Block Design subtests based upon the test records of 100 student nurses, 40 conscientious objectors, and 160 vocational guidance cases. The mean IQ of the sample group was 104.4. Rank orders of the subtest items were obtained by assigning the first position to that item which revealed the highest number of successfully passing subjects, and the last position to the item which showed the fewest number of successes. The "any success" method was utilized in determining item rankings. Minor but insignificant changes in item order were found on the Picture Arrangement, Block Design, and Similarities subtests. The authors recommended revising the Information, Comprehension, and Picture Completion subtest item orders because these revised orders varied considerably from the generally accepted order of presentation. They suggested that time and effort would be saved if the Picture Completion subtest were discontinued after three or four consecutive failures. An analysis of the Information, Comprehension, and Similarities subtests indicated that the transition of items from easy to hard was not gradual enough so that there were abrupt changes in item difficulty. Their results were later confirmed on one thousand psychiatric cases. No correlation coefficients were reported in this study, the data having been limited to a presentation of the revised item orders and the percentages of the sample that succeeded on various items. The

design of this investigation appears questionable in that the authors utilized the "any success" method in obtaining their revised subtest orders. It is the writer's contention that this method of determining item difficulty masks individual performances in that it fails to consider points earned for qualitatively superior responses. It is suspected that Rabin's criticism regarding the abrupt changes in item difficulty on the Comprehension and Similarities subtests stems in part from utilizing the "any success" method. For example, 85 per cent of his subjects passed items one through eight on the Comprehension subtest while 45 per cent and 65 per cent succeeded on items nine and ten, respectively. The abrupt change in difficulty from item eight to nine is indeed striking, but more important is the fact that 65 per cent of the sample passed all but one of the items, apparently, because passing was defined as any success. Also, the suggestion to discontinue Picture Completion items "after three or four consecutive failures" (17, p. 497) seems unwarranted in view of W-B I administration directions that, except for very low grade individuals, the entire series of pictures should be presented. (25, p. 181)

Jastak (11) made an item analysis of the W-B I in which he studied the order of item difficulty on each subtest. Sixteen hundred test records were obtained from hospital files and 1,172 of them were used in the study. The sample included 300 hospital attendants, 152 nurses, 400 patients of all categories, and 320 boys and girls aged eleven to sixteen who had been referred to the Delaware State Hospital. Jastak tabulated the item successes and computed the percent of success for each sample group as well as for the entire sample. An "all or none" system of scoring responses was used in reporting the percentage of successes because "...ordinal consistency of test

items is increased when partial and qualitative scores are disregarded in the final analysis." (11, p. 89) The items of each subtest were then ranked in the order of difficulty for each group and for the entire sample. The obtained ranks of each sample group were correlated with each other, with Wechsler's order, and with Jastak's revised order. The following correlations with Wechsler's rank orders were obtained by Jastak: Information .935; Comprehension .842; Arithmetic .976; Similarities .909; Vocabulary .951; Picture Arrangement 1.000; Picture Completion .768; Block Design .964; Digit Span 1.00; Object Assembly 1.00. The results of this study led Jastak to suggest that the order of items in the Picture Completion, Comprehension, Similarities, Information, and Vocabulary subtests of the W-B I should be changed. In contrast to Rapaport, Jastak suggested that an item analysis based on sex would give more informative results than one based on the various clinical categories.

Jastak's analysis is considerably more thorough than any of the previously cited studies in that his sample was larger and more heterogeneous. Also, he computed correlations indicating the degree of relationship between his revised subtest orders and those of Wechsler's. On the other hand, it appears that his "all or none" system of scoring responses, used in computing the percentage of successes, is theoretically questionable. Jastak points out examples where counting of partial and qualitative scores, i.e. any success, decreases the ordinal consistency of test items. It would seem however, that the determination of item difficulty and item arrangements on the basis of optimal performances fails to take into account scores that have actually been earned in individual situations. This writer would agree then an "all or

none" on an "any success," i.e. one or the other, system must be utilized in computing the item orders of the Picture Arrangement, Block Design, and Object Assembly subtests because maximal item scores vary within these subtests. Comprehension, Similarities, and Vocabulary subtest item responses, however, are scored zero, one, or two points depending on the quality of the subject's response. Hence, Jastak could have resolved the problem of scoring successes on the latter three subtests by computing an average score for each item.

Norman (15) studied the order of item difficulty in a young superior adult population, the majority of whom were college students. His findings were derived from the W-B I test records of 85 males and 68 females whose mean ages were 23.5 and 22.3 years, respectively. The Full Scale IQ statistics were: males, mean 127.5, standard deviation 5.2; females, mean 126.8 standard deviation 5.2. Over-all item success was computed by averaging male and female successes with each item. Although Norman's sample was much smaller in number and superior in intelligence to those of Jastak and Wechsler, all but one of his rank orders (Comprehension) correlated higher with Jastak's than with Wechsler's. Norman's correlation coefficients are presented in Table I. These coefficients are, of course, not directly comparable between Norman's superior adults and either Wechsler's or Jastak's group because Norman's sample was smaller (153) in number and superior in intelligence. Norman's thinking was that items causing difficulty in a higher intelligence group would generally cause difficulty in less intelligent groups. This hypothesis appears questionable when an analysis of Norman's revised orders is made. For example, the Information items, "H. Finn" and "Hamlet," were easier for his sample than "Pints" or "Height," suggesting that educated individuals of

TABLE I

CORRELATION COEFFICIENTS OBTAINED FROM A COMPARISON OF NORMAN'S ITEM ORDERS WITH THOSE OF JASTAK AND WECHSLER

Subtest	rho with Jastak	rho with Wechsler
Information	.855	.834
Comprehension	.479	.645
Similarities	.727	.671
Arithmetic	.936	.864
Vocabulary	.933	.917
Picture Arrangement	1.000	1.000
Picture Completion	.785	.517
Block Design	.786	.643

superior intelligence do encounter difficulty where less intelligent groups generally succeed.

Guertin (7) studied the effect of altered instructions and a revised item order on the Arithmetic subtest scores of 64 undergraduates, half of whom were males and half females. The experiment was designed to check the investigator's impression that resignation, which he describes as, "The emotional reaction to fear of failure resulting in giving up," can materially reduce the performance level on the Arithmetic subtest. Because of the high level of intelligence and arithmetic ability of the subjects, it was necessary to combine items from the W-B I and W-B II in their approximate order of difficulty to provide a sufficient number of difficult items at the ceiling of the scale. One-half of

the subjects, equally divided in sex, were administered the items with standard instruction; one-half were told, "Now I want you to do something else for me," followed immediately by the first problem. The 32 subjects in the standard instruction group did slightly poorer than those who received altered instructions, but the difference in mean scores was not significant at even the 5 per cent level of confidence. One half of the subjects were given the items in their usual order while the other half were presented with the most difficult problems (7-10) first. Contrary to the resignation hypothesis, the revised order of presentation produced a significantly higher mean score than did the conventional order. Guertin suggested that the latter results might be explained in terms of his sample, i.e. college students respond more effectively to intellectual challenges than, for example, subjects from a lower educational-occupational level. In spite of contradictory findings, Guertin thought it tenable to hold the hypothesis advanced. This investigation is of importance in that it indicated the effects of altered procedure upon subtest scores. Both experimental conditions, modifying instructions and revising the order of item presentation, resulted in increased mean raw scores. On the basis of these findings, it appears that any revision of the order of item presentation could conceivably raise test results spuriously in relation to the standardization population.

In an attempt to reduce the administration and scoring time for the W-B I Vocabulary subtest, Sorsky (23) introduced a revised method of item presentation. Using three consecutive item successes as a basal level and three successive failures as a point of cessation, he found that the middle 60 percent of the Vocabulary items discriminated as well as the administration of the full

range. He also reported a revised order of item difficulty for the Vocabulary subtest the data of which in unavailable. His revisions were expected to reduce scoring and administration time by about 50 percent.

Russell (20) used the W-B I test records of 200 consecutive neuropsychiatric patients in a military hospital as the basis for a revision of the Vocabulary subtest item order. The positions of many items were changed, e.g. "hari-kari" (34 in Wechsler's arrangement) was moved to seventeenth place in the revised order. Russell attributed his results to changes in the common vocabulary that have occurred since 1939.

Similar investigations of item order in the Revised Stanford-Binet Intelligence Scale have been reported. McNemar (13) reviewed the wide spread of individual test performances, and offered six reasons to account for the extreme variations in responses (1) item unreliability; (2) varying rates of mental growth; (3) a lack of steepness of differentiation among items; (4) sex differences; (5) personality disturbances; and (6) periodic variations in item difficulty. He stated that the performance of nearly all individuals must show scattering of passes and failures: "The order of difficulty of test items is not something that can be established once and for all. It can be expected to vary somewhat with samplings of different populations especially samplings of different countries." (12, p. 84)

Staniszewski (24) studied the effect of using a double ceiling in the administration of the Revised Stanford-Binet Scale. Twenty of her subjects achieved an average gain of 7.8 months in mental age when testing was extended beyond the first ceiling of failure. More than two-thirds (71 per cent) of the sample did not, however, make greater gains when administered items beyond the

usual point of termination. In view of the marked differences in difficulty which she found to exist between adjacent test (age) levels and between items within a single age level through the middle and upper ranges of the Stanford-Binet Scale, she suggested extending the range of testing beyond the first ceiling of failure. This suggestion was modified somewhat by that writer: "Yet, when the usual procedure of stopping with one ceiling is used by a well-trained examiner administering the scale to a child of apparent average ability, there is likelihood of a sizeable error in one of ten instances." (24, p. 24) Staniszewski's finding of marked differences in difficulty between adjacent test levels and between items within single age levels is less disconcerting to the reviewer when it is recalled that the Stanford-Binet Scale contains items which yield per cent passing with age scales that are none too steep and items that are not perfectly intercorrelated. (13, p. 71) Also, in constructing the Stanford-Binet no attempt was made to arrange items within a given age level in order of difficulty as was done on the Wechsler tests. Hence, there is more likelihood of performance variability on the Stanford-Binet. Staniszewski's suggestion to extend testing beyond the usual point of termination is acceptable because all successes and failures are taken account of in computing mental age on the Stanford-Binet.

In sum, the available literature revealed no published study dealing with or related to the ranking of subtest items in the WAIS. Several item analyses of the Wechsler-Bellevue Intelligence Scale, Form I, and similar studies of the Revised Stanford-Binet Intelligence Scale have been reported in the literature. The investigations of Altus (26), Rapaport (19), Rabin et al (17), Jastak (11), Norman (15), and Russell (20) revealed orders of item gradation which differed

from those recommended by Wechsler in his original and revised publications of the W-B I. Changes in the order of item presentation for the Arithmetic and Vocabulary subtests were reported by Guertin (7) and Sorsky (23), respectively. These studies, excepting Guertin's, revealed subtest orders of item difficulty that differed from the conventional subtest orders based on Wechsler's original standardization sample. In several cases the results were questioned because either the samples were not representative or the methods employed to determine revised orders were of doubtful propriety. McNemar's attempt (13) to account for scatter in the Stanford-Binet revealed six factors that might also explain variability of performance on the Wechsler tests. Staniszewski's results (24) indicated that in determining Stanford-Binet mental ages sizeable errors can be made when testing has been discontinued at the prescribed ceiling of failure. All of the above studies illustrated the importance with which item gradation is viewed in psychometric testing. It seems of value, therefore, to investigate the order of item arrangement on the WAIS.

CHAPTER III

PROCEDURE

According to Wechsler, an analysis of the difficulty of the WAIS items, based on the 1,700 cases in the standardization sample, revealed that the present item arrangement is essentially in order of least to most difficult. He commented that there were a few instances of inversion in order of difficulty but that these were of a minor sort (27, p. 29). The WAIS manual does not describe the procedure by which its final item orders were obtained. In view of Wechsler's comment that the WAIS is an extension and modification of the W-B I, it is presumed that the item orders of the WAIS subtests were established in the same manner as were those of its predecessor.

In constructing the W-B I subtests, Wechsler selected a large number of items and arranged them in their approximate order of difficulty. The items were then administered to experimental groups of known intelligence level. The items that discriminated poorly between different intelligence levels were excluded. Finally, the retained items were listed in an order of difficulty beginning with the one most frequently passed and ending with the item least often passed by the normative sample. Wechsler apparently anticipated variations in the difficulty of certain W-B I items: "The order in which questions are listed roughly approximates their order of difficulty for our sample of the population. No doubt, in different localities, the order will be somewhat different; it will also be affected to some extent by differences in

national origin." (26, p. 79)

A. Selection of the Test Records

The fifty test records used in this study were selected from a group of available WAIS tests that had been administered to one hundred young adults between May, 1956, and February, 1957. (27) The subjects had been selected according to the stratified sampling plan adopted by Wechsler and based on the 1950 United States census. (26, p. 5) The variables that governed this stratification were: age, sex, race, occupation, education, and religious affiliation. The original sample was reasonably representative of the eighteen and nineteen year old population with the exception that all of the subjects were residents of the Chicago area. The individuals used to fill the quotas were not randomly selected and they had to agree to be tested. These conditions were necessary, although they obviously do not conform to the theoretical definition of a random unit.

The test records selected were those of twenty-five men and twenty-five women whose mean age was computed to be eighteen years, eleven months. The range in age of the subjects was from 18.0 to 19.9 years, with a standard deviation of 6.9 months. Their mean number of years of education was 11.8, with a range of eight to fifteen years, and a standard deviation of 1.6 years. Of the fifty subjects, forty-nine were white and one was non-white. Thirteen were of Protestant faith, thirty-six were Roman Catholic, and one was Jewish.

The following is a division of the subjects based on the number of years of education completed at the time of testing. The percentages indicate the proportion of males and females in each category.

1. Completed eight years of formal education. (2% male, 2% female)

2. Completed nine to eleven years of formal education. (14% male, 8% female)
3. Completed twelve years of education. (26% male, 30% female)
4. Completed thirteen years or more of formal education. (8% male, 10% female)

An analysis of the WAIS records indicated that the mean Full Scale IQ for this sample was 109.26. The Full Scale IQ scores range from 88 to 131 with a standard deviation of 10.39. The somewhat higher than average mean IQ score probably stems from the requirement that all of the items of the subtests studied had been administered to each subject or his record was not selected for this study.

The fifty subjects whose records were selected came from various occupational settings. The following occupational categories were established by Wechsler on the basis of census groupings. (26, p. 7) The percentages following the occupational categories listed below indicate the proportion of the present sample group falling within each division.

1. Professional, technical, and kindred workers. (0% male, 4% female)
2. Farmers and farm managers. (2% male, 0% female)
3. Managers, officials, and proprietors except farm. (none)
4. Clerical, sales, and kindred workers. (8% male, 14% female)
5. Craftsmen, foremen, and kindred workers. (4% male, 0% female)
6. Operatives and kindred workers. (10% male, 6% female)
7. Private household workers. (none)
8. Service workers, except private household. (0% male, 6% female)
9. Farm laborers. (none)

10. Laborers. (4% male, 0% female)

11. Keeping house. (0% male, 8% female)

12. Students. (22% male, 12% female)

13. Others. (none)

All of the tests were administered and scored by three psychologists, including the writer, of the Chicago Catholic Charities Guidance Center. The testees' responses were recorded in full so that the scoring could be reviewed by another member of the department staff who is an experienced WAIS examiner. Scoring was in accord with the criteria listed in the WAIS manual.

As stated above, the fifty WAIS records used in this study were selected from one hundred tests that had been administered to an eighteen and nineteen year old sample considered representative of that age group. The entire sample was not selected because an examination of the records indicated that only fifty-three subjects had been administered all the items of the seven subtests studied. In order to determine the difficulty value of each item, therefore, it seemed necessary to select only the records of subjects who had been permitted to attempt all of the items. Also, in reviewing the studies described in Chapter II, it was noted, in most instances, that the authors failed to state whether or not every item had been administered to their subjects. Apparently, the per cent of the sample passing each item was calculated on the basis of the number of subjects that had attempted the item, i.e. subtests were discontinued when a subject failed the prescribed number of consecutive items. However valid that method may have been, a complete picture of item difficulty for a given sample cannot be known unless all of the subjects are permitted to attempt every item.

It was disconcerting to note that in selecting only the complete test records of the original sample some representativeness was necessarily lost. The greatest differences between the original and derived samples were in relation to race, intelligence, religion, and years of education. Fourteen per cent of the original sample was non-white, while only two per cent of the derived sample was non-white. The mean Full Scale IQ of the original sample was 101.90 as opposed to a mean IQ of 109.26 for the selected sample, the latter score closely approaching the lower extreme of the above-average range of intelligence. In the smaller sample seventy-two per cent of the subjects were Roman Catholics, twenty-six per cent were Protestant, and two per cent were Jewish as compared to the forty per cent Roman Catholic, fifty-two per cent Protestant, and six per cent Jewish individuals in the original sample. In years of education, seventy-four per cent of the smaller sample had completed at least twelve years of formal schooling while only forty-six per cent of the larger original sample had completed the same amount of education. This difference appears to stem in part from the proportion of students in the two samples: twenty-five per cent in the larger sample; thirty-four per cent in the smaller, derived sample. The reader is referred to Tables X and XI of the Appendix where comparisons between the two samples are presented. In general, a comparison indicated that there was only a moderate degree of resemblance between the groups. The derived, smaller sample was not considered representative of the United States eighteen to nineteen year old population. The results of this study, therefore, reflect the item difficulties encountered by a sample of eighteen and nineteen year old subjects the majority of whom were well-educated, somewhat more intelligent than average, Roman Catholic, white

individuals.

B. Determination of Item Orders and Correlations

When the Scoring had been checked, the following statistical steps were taken:

1. A point total for each item of the Information, Comprehension, Similarities, Arithmetic, Vocabulary, Block Design, and Picture Arrangement subtests[1] was obtained by adding the individual item scores.

2. The item orders of the subtests studied were determined in the following ways:

   a. The item order of the Information subtest was determined by dividing each item total by fifty (N=50) to obtain the percentage of the sample that succeeded on each item. The items were then arranged along a continuum of difficulty beginning with that item passed by the largest percentage of the subjects and terminating with the item passed by the smallest percentage of the sample. The "all or none" manner (one point or none) in which Information responses are scored obviates the difficulties encountered in ordering the items

---

1. The Picture Completion and Object Assembly subtests were not included in this study because every item of those scales is administered in testing. The Digit Span was excluded because the difficulty of its items depends on quantitative factors (the number of digits recalled) rather than item differences. The Digit Symbol subtest was not included because of the uniform nature of its structure, i.e. item repetition.

of subtests where partial credits or bonus points must be considered.

b. The item orders of the Comprehension, Similarities, and Vocabulary subtests were obtained by dividing the point totals for each item by fifty to arrive at an average or mean score for each item. The order of item difficulty for each of these three subtests was determined by the percentage of the sample that succeeded on each item. The item for which the highest average score was obtained was placed at the beginning of the series, followed in descending order by the remaining items and ending with that one for which the lowest average score was obtained, i.e. the hardest. The method of average scores was adopted for ordering the items of these subtests because of their scoring criteria, i.e. item responses are scored zero, one, or two points depending on the quality of the subject's verbalization. The writer believes that the exclusion of partial (one point) credits from consideration in the ordering of these items would fail to take into account credits which had actually been awarded in individual testings, and, hence, detract from a true picture of item difficulty. This position is contrary to that of Jastak's who stated, "The ordinal consistency of test items is increased when partial and qualitative scores are disregarded in the final analysis." (10, p. 89)

c. The Arithmetic, Block Design, and Picture Arrangement subtest item orders were determined on the basis of "any success." This method of obtaining rank orders was necessarily used because the maximal number of points that can be awarded varies by item within the three subtests. Also, the method of scoring Picture Arrangement responses, i.e. by the quality of successful response, introduced another variable that was removed by the procedure of tallying any and all successes. Hence, the item orders of these three subtests were obtained by arranging the items in a sequence beginning with that one for which the largest number of successes had been recorded and terminating with the item having the fewest number of successes.

3. When the writer's rank orders for each of the subtests studied had been arranged, they were correlated with Wechsler's item orders. The correlations were computed by the rank order coefficient of correlation (rho) method. The rank-difference correlation method was used because the sample was small and the data was in terms of rank orders rather than in terms of measurements. The following formula was used:

Rank-difference coefficient of correlation:

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

where D = difference in rank in the two series

ΣD^2 = sum of squares of all such differences

N = number of paired cases (items)

According to Garrett, coefficients of correlation calculated from a few cases are not very reliable, and their value lies in suggesting the possible existence of relationship. He states, "Since rho is at best only an approximate measure of relationship..., it is hardly worth while computing its standard error." (5, p. 355) On the other hand, Guilford (28, p. 313) considers an obtained rho as an approximation of the Pearson product-moment coefficient because the difference between rho and "r" is never greater than .018, "r" always being greater than rho. The reliabilities of the obtained rhos were estimated by the following formula, suggested by Guilford:

Standard error of a rho coefficient:

$$\sigma_\rho = \frac{1.04\,(1-\rho^2)}{N-1}$$

where ρ^2 = the obtained rho, squared

N = number of cases (items)

CHAPTER IV

RESULTS

In the Information subtest, where testing may be terminated after five consecutive item failures, there were five subjects who, having failed five successive items, succeeded one or more times when administered the entire sequence of items. The following items were relatively easier for the subjects used in this study than for those of Wechsler's standardization sample: "Hamlet" (our rank order position 7, Wechsler's 15); "Vatican" (ours 12, Wechsler's 16); "Senators" (ours 15, Wechsler's 21); "Iliad" (ours 18, Wechsler's 24); "Koran" (ours 21, Wechsler's 26). Among the items found more difficult for the subjects were: "Italy" (ours 16, Wechsler's 12); "Clothes" (ours 17, Wechsler's 13); "Paris" (ours 24, Wechsler's 17); and "Population" (ours 26, Wechsler's 20). The first four and last three items remain in the positions assigned them by Wechsler. In general, over one-half of the subjects succeeded on twenty-five of the twenty-nine items. This frequency of success appears to stem from two factors. In the first place, thirty-four per cent of the sample were students and seventy-four per cent of the subjects had completed twelve or more years of formal schooling. Only twenty-nine per cent of Wechsler's eighteen and nineteen year old standardization population were students, and thirty-one per cent of his subjects in that age group had completed twelve years of school. Secondly, the mean IQ for our sample (109.26) reflects a somewhat higher than average level of intelligence. Apparently ed-

ucational and intellectual advantages assisted the subjects of this study in their performance on the Information subtest. The greater ease with which "Vatican" and "Koran" items were passed appears to be a function of religious affiliation in that seventy-two per cent of the subjects are Roman Catholic. Presumably, they received religious instruction in school that included highlights of other faiths, as well as of their own. Our revised ordering of items correlated .92 with that of Wechsler's. This estimate was significant at the one per cent level of confidence, and its standard error was .04. The item order of the Information subtest is presented in Table II.

On the Comprehension subtest, where testing may be terminated after four consecutive item failures, there was only one subject who after failing four successive items proceeded to earn additional credits when administered the entire series. The following items were relatively easier for the subjects of this study than for Wechsler's sample: "Forest" (ours 6, Wechsler's 9); "Deaf" (ours 8, Wechsler's 10); "Marriage" (ours 9, Wechsler's 12); and "Brooks" (ours 11, Wechsler's 13). The items found more difficult by the subjects were: "Taxes" (ours 7, Wechsler's 6); and "City Land" (ours 13, Wechsler's 11); "Iron" (ours 10, Wechsler's 7); and "Child Labor" (ours 12, Wechsler's 8). There is, however, perfect agreement between our results and Wechsler's arrangement of items one, two, three, four, five, and fourteen. Only item eight, "Child Labor," appears significantly misplaced, having been considerably more difficult for our subjects than its position would suggest. An examination of scoring criteria and sample answers revealed that "Child Labor" is the only comprehension item that requires two acceptable answers for a perfect score, i.e. two points. It is obviously difficult to earn two points on this item, "Why are

child labor laws needed?", when the question does not specifically ask for two or more reasons. Our revised order correlated .88 with Wechsler's. This estimate was significant at the one per cent level of confidence, and its standard error was .07. The item order of the Comprehension subtest is presented in Table III.

A high degree of correspondence was found to exist between our item order and that of Wechsler's for the Arithmetic subtest. Only problem six appears misplaced, being slightly more difficult for the subjects of this study than its ordinal position suggests; item six became number eight in our arrangement. Items one through ten appear to lack a graduated degree of steepness in difficulty since eighty-four per cent of the subjects passed them. The ease with which the first ten Arithmetic problems were solved might be accounted for in terms of the subjects' educational and intellectual advantages previously cited in connection with their performance on the Information subtest. Also, the preponderence of successes on items eleven through fourteen, the time credit problems, indicates that for our subjects the Arithmetic test lacks an adequate ceiling. One subject passed an item after failing four consecutive problems, the prescribed ceiling for discontinuance of items. The coefficient of correlation between our item ranking and that of Wechsler's was .99, a degree of relationship significant at the one per cent level of confidence. The standard error of the correlation coefficient was .01. The item order of the WAIS Arithmetic Subtest is presented in Table IV.

In the Similarities subtest there was considerable variation between Wechsler's rank order and our results, although there was no case in which a subject passed an item after four consecutive failures. The following items

TABLE II

ITEM ORDER OF THE WAIS INFORMATION SUBTEST

N=50)

	Item	Number of Successes	Percentage Succeeding	Revised Order
1.	Flag	50	100	1
2.	Ball	50	100	2
3.	Months	50	100	3
4.	Thermometer	50	100	4
5.	Rubber	49	98	6
6.	Presidents	46	92	9
7.	Longfellow	50	100	5
8.	Weeks	46	92	10
9.	Panama	45	90	11
10.	Brazil	47	94	8
11.	Height	43	86	13
12.	Italy	34	68	16
13.	Clothes	34	68	17
14.	Washington	40	80	14
15.	Hamlet	49	98	7
16.	Vatican	44	88	12
17.	Paris	15	30	24
18.	Egypt	22	44	20
19.	Yeast	20	40	22
20.	Population	12	24	26
21.	Senators	38	76	15
22.	Genesis	30	60	19
23.	Temperature	20	40	23
24.	Iliad	34	68	18
25.	Blood Vessels	15	30	25
26.	Koran	22	44	21
27.	Faust	4	8	27
28.	Ethnology	1	2	28
29.	Apocrypha	1	2	29

Rho between Wechsler's order and revised order: .92

TABLE III

ITEM ORDER OF THE WAIS COMPREHENSION SUBTEST

	Point Totals	Average Score	Revised Rank Order
1. Clothes	100	2.00	1
2. Engine	100	2.00	2
3. Envelope	100	2.00	3
4. Bad Company	93	1.86	4
5. Movies	78	1.56	5
6. Taxes	74	1.48	7
7. Iron	63	1.26	10
8. Child Labor	59	1.18	12
9. Forest	77	1.54	6
10. Deaf	69	1.38	8
11. City Land	55	1.10	13
12. Marriage	67	1.34	9
13. Brooks	60	1.20	11
14. Swallow	39	0.78	14

Rho between Wechsler's order and revised order: .88

TABLE IV

ITEM ORDER OF THE WAIS ARITHMETIC SUBTEST

Item	Number of Successes	Percentage of N Succeeding	Revised Rank Order
1.	50	100	1
2.	50	100	2
3.	50	100	3
4.	49	98	4
5.	47	94	5
6.	43	86	8
7.	47	94	6
8.	46	92	7
9.	42	84	9
10.	42	84	10
11.	36	72	11
12.	30	60	12
13.	18	36	13
14.	11	22	14

Rho between Wechsler's order and revised order: .99

were more difficult for the subjects than expected on the basis of their ordinal positions: "Orange-Banana" (our rank 4, Wechsler's 1); "Axe-Saw" (ours 5, Wechsler's 3); and "Air-Water" (ours 10, Wechsler's 7). The following items were easier than expected: "North-West" (ours 1, Wechsler's 5); "Dog-Lion" (ours 2, Wechsler's 4); "Table-Chair" (ours 6, Wechsler's 8); and "Fly-Tree" (ours 11, Wechsler's 13.) One would suspect that the "Orange-Banana" item was more difficult because it is the initial item in a subtest that requires a considerably different mental approach than the previously administered Information, Comprehension, and Arithmetic subtests. This writer has noted in individual testing situations that subjects are frequently confused by the introductory question, "In what way are an Orange and a Banana alike?" Occasionally they respond too quickly to permit an adequate evaluation of the problem posed. It would seem that the W-B I instructions, "I am going to name two things which are the same or alike in certain ways and I want you to tell me in what way they are alike...," introduced the Similarities subtest more understandably than the WAIS instructions. The "North-West" item was easier for our subjects, probably, because as residents of the Chicago area they are accustomed to visualizing locations in terms of their position from Lake Michigan. Our revised ordering correlated .84 with Wechsler's presentation. This estimate was significant at the one per cent level of confidence, and its standard error was .09. The item order of the WAIS Similarities Subtest is presented in Table V.

An analysis of the Vocabulary subtest item arrangement arrived at in this investigation indicated that several words are significantly misplaced in the regular item order. Items nine, twelve, twenty-four, twenty-six, and thirty-

TABLE V

ITEM ORDER OF THE WAIS SIMILARITIES SUBTEST

(N=50)

		Point Totals	Average Score	Revised Rank Order
1.	Orange-Banana	89	1.78	4
2.	Coat-Dress	92	1.84	3
3.	Axe-Saw	86	1.72	5
4.	Dog-Lion	96	1.92	2
5.	North-West	97	1.94	1
6.	Eye-Ear	74	1.48	7
7.	Air-Water	39	0.78	10
8.	Table-Chair	77	1.54	6
9.	Egg-Seed	52	1.04	9
10.	Poem-Statue	56	1.12	8
11.	Wood-Alcohol	16	0.32	12
12.	Praise-Punish	15	0.30	13
13.	Fly-Tree	34	0.68	11

Rho between Wechsler's order and revised order: .84

three of Wechsler's order became items four, five, sixteen, eighteen and twenty-six respectively in the revised order, a positional difference of at least five items in each case. The greater ease with which the above items were passed by the sample group is of importance when considering that adminis-

tration procedure for this subtest permits the examiner to terminate testing when five consecutive failures have been recorded. Six subjects of the sample group obtained higher scores when administered all of the Vocabulary terms. In each of these cases five consecutive failures had been recorded.

The following items of the Vocabulary test were found more difficult by the subjects of this study than was expected on the basis of Wechsler's arrangement: "Winter" (our ranking 9, Wechsler's 4); "Fabric" (ours 11, Wechsler's 7); "Slice" (ours 25, Wechsler's 8); "Commence" (ours 23, Wechsler's 15); and "Remorse" (ours 32, Wechsler's 23). The most significantly misplaced item, "Slice," was more difficult because scoring criteria requires a subject to include the idea of thinness in his definition in order to gain a maximal score of two points. The subjects of this study, however, limited their responses to the cutting aspect of "Slice" without mentioning the accompanying, required thinness. Hence, poverty of content lowered the majority of the subjects' scores on that item. Another item, "Commence," was difficult for the subjects because they frequently confused it with "Commencement." The last seven items, thirty-four through forty, varied only slightly from the positions Wechsler has them in order of difficulty. The revised Vocabulary order correlated .92 with Wechsler's, significant at the one per cent level. The standard error of the correlation coefficient was .03. In spite of the high coefficient between our revised order and Wechsler's presentation, the Vocabulary test seems to be poorly structured for the clinical analysis of intratest scatter. The largest variations were found between the fourth and thirty-third words. The item order of the WAIS Vocabulary Subtest is presented in Table VI.

TABLE VI

ITEM ORDER OF THE WAIS VOCABULARY SUBTEST

(N=50)

	Item	Point Totals	Average Score	Revised Rank Order
1.	Bed	100	2.00	1
2.	Ship	100	2.00	2
3.	Penny	100	2.00	3
4.	Winter	92	1.84	9
5.	Repair	95	1.90	8
6.	Breakfast	98	1.96	6
7.	Fabric	86	1.72	11
8.	Slice	63	1.26	25
9.	Assemble	100	2.00	4
10.	Conceal	98	1.96	7
11.	Enormous	87	1.74	10
12.	Hasten	99	1.98	5
13.	Sentence	78	1.56	17
14.	Regulate	82	1.64	14
15.	Commence	65	1.30	23
16.	Ponder	83	1.66	12
17.	Cavern	73	1.46	19
18.	Designate	81	1.62	15
19.	Domestic	83	1.66	13
20.	Consume	67	1.34	21
21.	Terminate	66	1.32	22
22.	Obstruct	72	1.44	30
23.	Remorse	43	0.86	32
24.	Sanctuary	81	1.62	16
25.	Matchless	64	1.28	24
26.	Reluctant	77	1.54	18
27.	Calamity	60	1.20	27
28.	Fortitude	48	0.96	30
29.	Tranquil	46	0.92	31
30.	Edifice	30	0.60	33
31.	Compassion	56	1.12	28
32.	Tangible	52	1.04	29
33.	Perimeter	62	1.24	26
34.	Audacious	29	0.58	34
35.	Ominous	16	0.32	36
36.	Tirade	12	0.24	39
37.	Encumber	20	0.40	35

TABLE VI (Continued)

ITEM ORDER OF THE WAIS VOCABULARY SUBTEST

(N=50)

Item	Point Totals	Average Score	Revised Rank Order
38. Plagarize	16	0.32	37
39. Impale	16	0.32	38
40. Travesty	2	0.04	40

Rho between Wechsler's order and revised order: .92

An almost complete agreement was found to exist between our rank order and Wechsler's arrangement of the Block Design subtest items. The single discrepancy found, a reversal of items two and three, stems from a singular failure on item two whereas all of the subjects of this study passed item three. Also, there was no occasion in which an individual failed three consecutive items, the prescribed criteria for discontinuance of this subtest, and then went on to pass an additional item. In view of the fact that ninety per cent of the sample passed (any success) the first seven items, the Block Design subtest appears to lack an adequate ceiling for our sample of the 18-19 year age group, although it is somewhat select in that the mean IQ of the subjects was significantly superior to that of the standardization sample. The final rank order correlated .99 with Wechsler's arrangement. This figure is significant at the one per cent level of confidence; its standard error was .01. The item order of the WAIS Block Design subtest is presented in Table VII.

TABLE VII

ITEM ORDER OF THE WAIS BLOCK DESIGN SUBTEST

(N=50)

Item	Point Totals*	Number of Successes	Percentage of N Succeeding	Revised Rank Order
1.	200	50	100	1
2.	190	49	98	3
3.	200	50	100	2
4.	196	49	98	4
5.	196	49	98	5
6.	180	45	90	6
7.	232	45	90	7
8.	146	30	60	8
9.	130	25	50	9
10.	115	23	46	10

Rho between Wechsler's order and revised order: .99

* Items 1, 2, 7, 8, 9, and 10 include trial and time credits.

On the Picture Arrangement subtest, where discontinuance of items is authorized when a subject fails both trials of the first two items, it was found that all of the subjects had passed the first two items. Item six, "Flirt," was somewhat easier for the subjects of this study (our rank 3) than for Wechsler's sample. The greater ease with which the "Flirt" pictures were arranged (only two subjects failed this item) appears to reflect the subjects'

heightened awareness, or set, for the establishment of heterosexual relationships. The "Flirt" pictures might be more adequately renamed "Pick up." The items found slightly more difficult were: "Louie" (our rank 5, Wechsler's 4); and "Enter" (ours 6, Wechsler's 5). The rank order based on our results correlated .86 with that of Wechsler's. This figure was significant at the one per cent level of confidence, the standard error of the estimate having been calculated as .10. The item order of the WAIS Picture arrangement subtest is presented in Table VIII.

TABLE VIII

ITEM ORDER OF THE WAIS PICTURE ARRANGEMENT SUBTEST
(N=50)

	Item	Point Totals*	Number of Successes	Percentage of N Succeeding	Revised Rank Order
1.	Nest	198	50	100	1
2.	House	200	50	100	2
3.	Hold Up	188	47	94	4
4.	Louie	176	44	88	5
5.	Enter	148	37	74	6
6.	Flirt	128	48	96	3
7.	Fish	133	29	58	7
8.	Taxi	116	28	56	8

Rho between Wechsler's order and revised order: .86

* Point totals of items 1, 2, 6, 7, and 8 include time credits.

CHAPTER V

SUMMARY AND CONCLUSIONS

The purpose of this investigation was to determine whether the arrangement of the Information, Comprehension, Arithmetic, Similarities, Vocabulary, Block Design, and Picture Arrangement subtest items of the WAIS is in order of least to most difficult for an eighteen and nineteen year old sample of the population. The available literature revealed no published study dealing with or related to the ranking of subtest items in the WAIS. Several item analyses of the Wechsler-Bellevue Intelligence Scale, Form I, and similar studies of the Revised Stanford-Binet Intelligence Scale that have been reported in the literature were reviewed. The investigations of Altus (26), Rapaport (19), Rabin et al (17), Jastak (11), Norman (15), and Russell (20), revealed orders of item gradation which differed from those recommended by Wechsler in his original and revised publications of the W-B I. Changes in the order of item presentation for the Arithmetic and Vocabulary subtests were reported by Guertin (7) and Sorsky (23), respectively. These studies, excepting Guertin's, revealed subtest orders of item difficulty that differed from the conventional subtest orders based on Wechsler's original standardization sample. In several cases the results were questioned because either the samples were not representative or the methods employed to determine revised orders were of doubtful propriety. McNemar's attempt (13) to account for scatter in the Stanford-Binet revealed six factors that might also explain

variability of performance on the WAIS. Staniszewski's results (24) indicated that sizeable errors can be made in determining Stanford-Binet mental ages when testing has been discontinued at the prescribed ceiling of failure. These studies illustrated the importance with which item gradation is viewed in psychometric testing.

The fifty WAIS records used in this study were selected from one hundred tests that had been administered to an eighteen and nineteen year old sample considered representative of that age group in terms of education, occupation, religious affiliation, and sex. The entire sample was not selected because an examination of the records indicated that only fifty-three subjects had been administered all the items of the seven subtests studied. In selecting only the complete records of the original sample, some representativeness was lost. The derived, smaller sample used in this study was not considered representative of the eighteen and nineteen year old population. The results of this study, therefore, reflect the item difficulties encountered by a select sample the majority of whom were well-educated, somewhat more intelligent than average, Roman Catholic, white individuals.

The WAIS records were analysed to ascertain the order of difficulty of the items in each of the subtests studied. The items were arranged on a continuum beginning with that one for which the highest average score or the highest percentage of success was computed and terminating with the item for which the lowest average score or lowest percentage of success was obtained. The degree of relationship between the revised item orders and the prescribed WAIS item orders was computed by the rank order coefficient of correlation (rho) method.

Several important differences were found to exist between the WAIS item orders and those derived from the test records used in this study. Considerable discrepancy between Wechsler's arrangement and the revised ordering of the Similarities subtest items was found. These orders correlated .84. Other subtest item orders found to vary somewhat from those of Wechsler's were: Picture Arrangement (rho=.86), and Comprehension (rho=.88). None of the subjects who failed the prescribed number of consecutive items necessary for discontinuance of the Similarities and Picture Arrangement subtests, proceeded to gain additional credits when administered the entire series. The items of these two subtests were, therefore, arranged accurately enough in order of difficulty for valid results. There was one subject who after failing four consecutive Comprehension items proceeded to gain an additional credit when administered the entire sequence of items. In general, greater ease in testing and economy of time would have been expected if the items of the Similarities, Picture Arrangement, and Comprehension subtests had been administered in their revised order.

A high degree of correspondence was found to exist between Wechsler's arrangement of the Block Design and Arithmetic subtest items and the item orders for those subtests resulting from this analysis. The correlation coefficients obtained were .99 and .99, respectively.

In spite of the high coefficients found between the revised orders and Wechsler's arrangement of the Vocabulary (.92) and Information (.92) items, these subtests appear to be poorly structured for administration to eighteen and nineteen year old testees. Seven of the subjects gained additional credits when administered all of the Vocabulary items; six increased their

scores when administered all of the Information items. In each of the above cases, the prescribed number of consecutive failures for subtest discontinuance had been recorded. On the basis of these findings, it appears necessary to administer Information and Vocabulary items beyond the prescribed ceiling if absolute scores are desired. It should be reemphasized, however, that the select character of the sample undoubtedly influenced the results although the extent of influence is unknown.

The writer suggests that a more extensive analysis based on a sample representative of the most frequently tested age groups be made so as to more accurately ascertain the difficulty value of the WAIS subtest items. If the results conflict appreciably with the WAIS item orders, it may be advisable to revise the present item orders of several subtests. In view of the non-representative character of the sample and the consequent limitations of this study, the writer would recommend administering all of the Vocabulary and Information subtest items to eighteen and nineteen year old subjects of apparently average or above average intelligence. Overtesting appears to be a safer procedure than immediate discontinuance of testing when a certain number of consecutive failures constitutes a basis for stopping. It should be noted, however, that administering all of the items might raise an individual IQ spuriously in relation to the standardization population. Also, administering all of these items would notably increase testing time. One value of this study was that it pointed out discrepancies in the arrangement of WAIS subtest items that are presumed to be in order of least to most difficult. Examiners should be cautious in making clinical interpretations based on the intra-test variability of an individual's performance on the WAIS. The results of this

investigation revealed, with minor exceptions, that the present arrangement of WAIS items in the subtests studied is essentially in order of least to most difficult.

BIBLIOGRAPHY

1. Altus, W. D. "The Differential Validity and Difficulty of Subtests of the Wechsler Mental Ability Scale." Psychol. Bull., 1945, 42, 238-Bull., 1945, 42, 238-249.

2. Brown, M. H., and Bryan, G. E. "Sex Differences in Intelligence." J. Clin. Psychol., 1955, 11, 303-304.

3. Cole, D., and Weleba, L. "Comparison Data on the Wechsler-Bellevue and the Wechsler Adult Intelligence Scale." J. Clin. Psychol., 1956, 12, 198-199.

4. French, E. G., and Hunt, W. A. "The Relationship of Scatter in Test Performance to Intelligence Level." J. Clin. Psychol., 1951, 7, 95-98.

5. Garrett, H. E. Statistics in Psychology and Education. (4th Ed.) New York: Longman's, Green Co., 1953.

6. Goolishian, H. A., and Ramsay, R. "The Wechsler-Bellevue, Form I, and the Wechsler Adult Intelligence Scale: A Comparison." J. Clin. Psychol., 1956, 12, 147-151.

7. Guertin, W. H. "The Effect of Instruction and Item Order on the Arithmetic Subtest of the Wechsler-Bellevue." J. Genet. Psychol., 1954, 85, 79-83.

8. Guertin, W. H., Frank, C. H., and Rabin, A. I. "Research with the Wechsler-Bellevue Intelligence Scale: 1950-55." Psychol. Bull., 1956, 53, 235-257.

9. Guilford, J. P. Fundamental Statistics in Psychology and Education. (2nd Ed.) New York: McGraw-Hill Book Co., 1950.

10. Gurvitz, M. S. "Some Defects of the Wechsler-Bellevue." J. Consult. Psychol., 1952, 16, 124-126.

11. Jastak, J. "An Item Analysis of the Wechsler-Bellevue Tests." J. Consult Psychol., 1950, 14, 88-94.

12. Jastak, J. "Problems of Psychometric Scatter Analysis." Psychol. Bull., 1949, 46, 177-197.

13. McNemar, Q. The Revision of the Stanford-Binet Scale. Boston: Houghton Mifflin, 1942.

14. Mech, E. "Item Analysis and Discriminative Value of Selected Wechsler-Bellevue Subtests." J. Educ. Res., 1953, 47, 241-260.

15. Norman, R. D. "Sex Differences and Other Aspects of Young Superior Adult Performance on the Wechsler-Bellevue." J. Consult. Psychol., 1953, 17, 411-418.

16. Rabin, A. I. "The Use of the Wechsler-Bellevue Scales with Normal and Abnormal Persons." Psychol. Bull., 1945, 42, 410-422.

17. Rabin, A. I., Davis, J. C., and Sanderson, M. H. "Item Difficulty on Some Wechsler-Bellevue Subtests." J. Appl. Psychol., 1946, 30, 493-500.

18. Rabin, A. I., and Guertin, W. H. "Research with the Wechsler-Bellevue Test: 1945-1950." Psychol. Bull., 1951, 48, 211-248.

19. Rapaport, D. Diagnostic Psychological Testing. Chicago: Year Book Publishers, 1945, Vol. I.

20. Russell, G. E. "Wechsler-Bellevue Vocabulary Subtest Items: Revised Order of Words." Med. Tech. Bull., 1954, 5, 143-148.

21. Schnadt, F. "Certain Aspects of Wechsler-Bellevue Scatter at Low IQ Levels." J. Consult. Psychol., 1952, 16, 456-461.

22. Schofield, W. "Critique of Scatter and Profile Analysis of Psychometric Data." J. Clin. Psychol., 1952, 8, 16-22.

23. Sorsky, F. "A Briefer Method for the Wechsler-Bellevue Vocabulary Subtest" Unpublished Master's Thesis, University of Pittsburg, 1949.

24. Staniszewski, Irene. "The Effect of a Double Ceiling in the Administration of the Revised Stanford-Binet Scale." Unpublished Master's Thesis, Loyola University, Chicago, Illinois, 1948.

25. Strange, F. B., and Palmer, J. O. "A Note on Sex Differences on the Wechsler-Bellevue Tests." J. Clin. Psychol., 1953, 9, 85-87.

26. Wechsler, D. The Measurement of Adult Intelligence. (3rd Ed.) Baltimore: Williams and Wilkins, 1944.

27. Wechsler, D. Manual for the Wechsler Adult Intelligence Scale. New York: Psychological Corporation, 1955.

28. Whelan, F. D. "A Study of the Reliability of the Wechsler Adult

Intelligence Scale with Eighteen and Nineteen Year Olds." Unpublished Doctoral Dissertation, Loyola University, Chicago, Illinois, 1957.

APPENDIX I

TABLE IX

SUMMARY INDICATING THE NUMBER OF ITEMS CARRIED OVER WITH NO MORE THAN MINOR CHANGES FROM THE W-B I TO THE WAIS (WECHSLER)

Test	Number of Items in W-B I	Number of Items Retained from W-B I	Number of Items in WAIS
Information	26	16	29
Comprehension	12	8	14
Arithmetic	10	5	14
Similarities	12	10	13
Digit Span - Forward	7	7	7
- Backward	7	7	7
Vocabulary	42	0	40
Digit Symbol	67	67	90
Picture Completion	15	11	21
Block Design	9	7	10
(Although 7 designs have been retained from W-B I, the blocks for all designs in WAIS are only red and white, the yellow and blue of the W-B I blocks having been eliminated.)			
Picture Arrangement	7	6	8
Object Assembly	3	3	4
(The Profile assembly has been reduced in size.)			

APPENDIX II

TABLE X

PER CENT OF SUBJECTS ACCORDING TO OCCUPATIONAL CATEGORY

Occupational Category	Male N=25		Female N=25	
	Percentage in Original Sample	Percentage in Sample	Percentage in Original Sample	Percentage in Sample
Professional, technical, kindred workers	2	0	4	8
Managers, proprietors, except farm	4	0	0	0
Clerical, sales, and kindred workers	12	16	26	28
Craftsmen, foremen, and kindred workers	8	8	0	0
Operatives and kindred workers	16	20	6	12
Private household workers	0	0	4	0
Service workers, except private household	4	0	6	12
Laborers, except farm	22	8	0	0
Keeping House	0	0	30	16
Students	28	44	22	24
Farmers and farm managers	0	4	0	0
Others - unable to work, voluntarily idle	4	0	2	0

APPENDIX III

TABLE XI

PER CENT OF SUBJECTS ACCORDING TO EDUCATIONAL LEVEL

Years of School Completed	Male N=25		Female N=25	
	Percentage in Original Sample	Percentage in Sample	Percentage in Original Sample	Percentage in Sample
8 or less years	26	4	14	4
9 - 11 years	38	28	30	16
12 years	28	52	48	60
13 - 15 years	8	16	8	20

APPROVAL SHEET

The thesis submitted by William G. Klett, has been read and approved by three members of the Department of Psychology.

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated, and that the thesis is now given final approval with reference to content, form, and mechanical accuracy.

The thesis is therefore accepted in partial fulfillment of the requirements for the Degree of Master of Arts.

April 10, 1958
Date

Frank J. Kobler
Signature of Adviser